

# Rerouting an Organocatalytic Reaction by Intercepting its Reactive Intermediates

Santosh C. Gadekar<sup>‡</sup>, Vasudevan Dhayalan<sup>‡</sup>, Ashim Nandi, Inbal L. Zak, Shahar Barkai, Meital Shema Mizrachi, Sebastian Kozuch and Anat Milo\*

Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

## Supporting Information

### Table of Contents

Mathematical Modelling.....	2
Linear Regression .....	3
Multinomial Logistic Regression .....	8
Substrate Key for Multinomial Logistic Regression.....	10
Published Model - Results .....	19
DFT Calculations.....	22
Computational method.....	22
Mechanism and energy profile.....	22
Alcohols as H/D source .....	25
Energies of all the species involved in the NHC-catalyzed HDE process.....	32
References.....	35

## Mathematical Modelling

Geometry optimizations and frequency calculations were carried out using Gaussian 16 software.<sup>1</sup> The functional used for DFT calculation is M06-2X, which was previously benchmarked for thermodynamic and kinetic accuracy of main group elements, and for non-covalent interactions.<sup>2,3</sup> A triple zeta potential basis-set (def2-TZVP)<sup>4,5</sup> was chosen based on Zhao and Truhlar's evaluation of the M06-2X functional for organic molecules, indicating that a triple zeta quality is generally more quantitative.<sup>5</sup> As this study seeks correlations, we opted not to incorporate scaling factors for vibrational terms.<sup>6</sup> This simplification was justified by the assumption that a constant scaling factor would neither change the descriptive parameters identified, nor the relationship between them. Sterimol values were calculated for the geometry optimized structures using the open source Python module Sterimol.<sup>7</sup> Sterimol values<sup>8</sup> were calculated along the substituent axis such that the principal axis begins on the carbonyl carbon. Thus, L is the added length of the substituted ring and the carbonyl – ring bond distance, B1 is the minimal width perpendicular to the principal axis L and B5 is the maximal width perpendicular to the principal axis L (for the most studied aromatic aldehydes, B1 and B5 correspond the thickness and the width of the substituted ring respectively - see figure below).

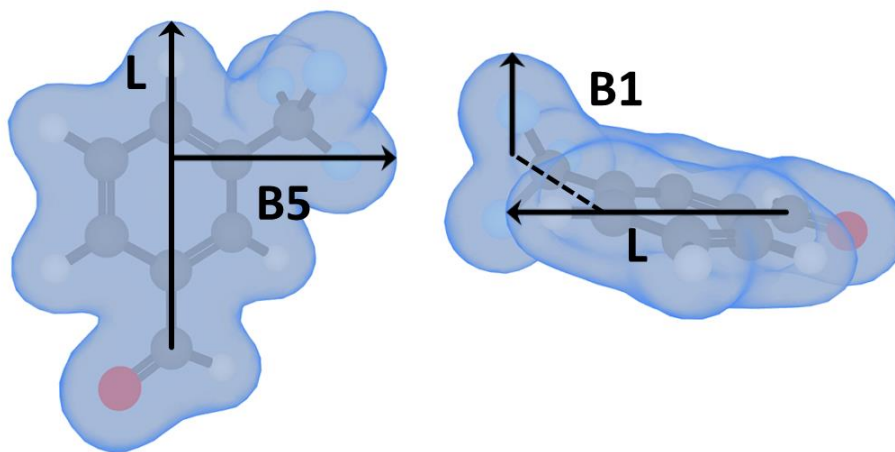


Figure 1. Verloop's Sterimol parameters L, B1 and B5 demonstrated upon 2-Fluorobenzaldehyde.

NPA charges were calculated using the NBO 3.1 extension incorporated in Gaussian.<sup>9</sup> Dipole moments for multinomial logistic regression were calculated for the geometry optimized structures based on NPA charges with respect to a consistent Cartesian origin

located on the carbonyl carbon with an orientation such that the carbonyl's carbon-oxygen bond acts as the y axis and the x axis shares a plane with the substituted ring. In the linear regression models, the origin is set at the center of the benzene ring with the y axis towards the carbonyl-ring bond and the x axis in the plane of the ring (see figure below).

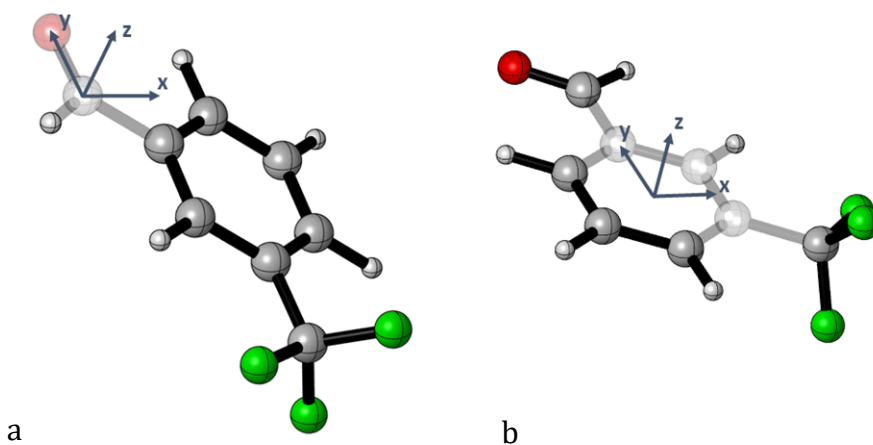


Figure 2. Graphic representation of the two coordinate systems used for extracting dipole moments in mathematical modelling: a. multinomial logistic regression (in figure 5 of the main text). b. linear regression (in figure 3 of the main text).

## Linear Regression

Similarly, to previous work<sup>10-14</sup> model development involved 4 stepwise regression algorithms that assess the significance of each parameter by applying statistical criteria. In order to realize this assessment, each set of parameter values was normalized by subtracting its respective mean and dividing by its standard deviation. The 4 stepwise regression algorithms are built in MATLAB<sup>15</sup> (R2019b, version 9.7.0.1261785) statistical toolbox and add or remove normalized parameters from an initial model according to a p-value threshold.<sup>16</sup> Additional suggestions for models were inputted manually based on the results of the 4 preliminary algorithm results or based on our mechanistic hypotheses. A linear fit was performed to probe each manually suggested model's statistical probability, and each suggestion was examined as an initial model for stepwise regression to seek a more effective model. We validated the models with a leave-one-out validation included in our MATLAB code and using a modified function published by Guo et al. that separates the set into a test and validation set, predicts the validations and reports the goodness-of-fit  $Q^2$

value.<sup>17</sup> In both models a 3-fold validation was performed, the procedure was repeated over 500 iterations and an average  $Q^2$  value was reported. The code together with the correlation and validation functions are provided in separate files along with the complete data set in an Excel format. The modelling source code together with optimized structures (in .xyz format) used to compute the parameters is provided on our GitHub page.

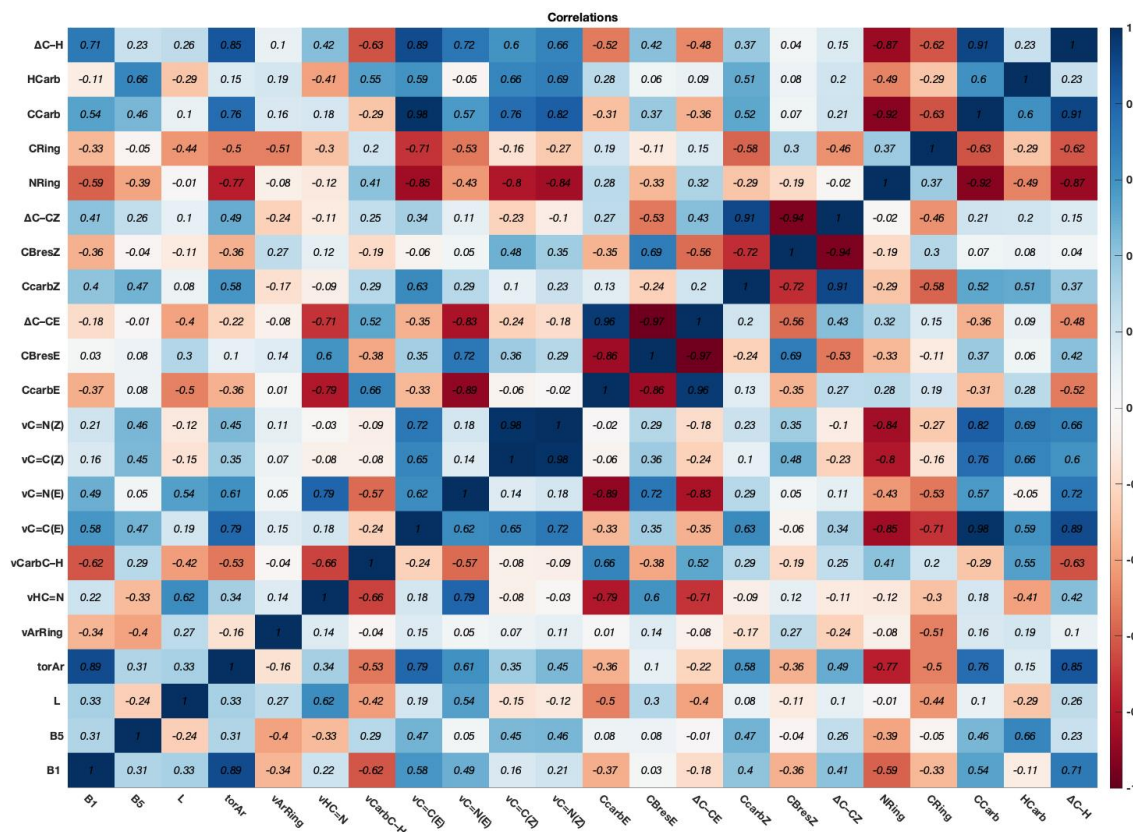


Figure 3. Correlation matrix for log(rate) model (in figure 3 of the main text).

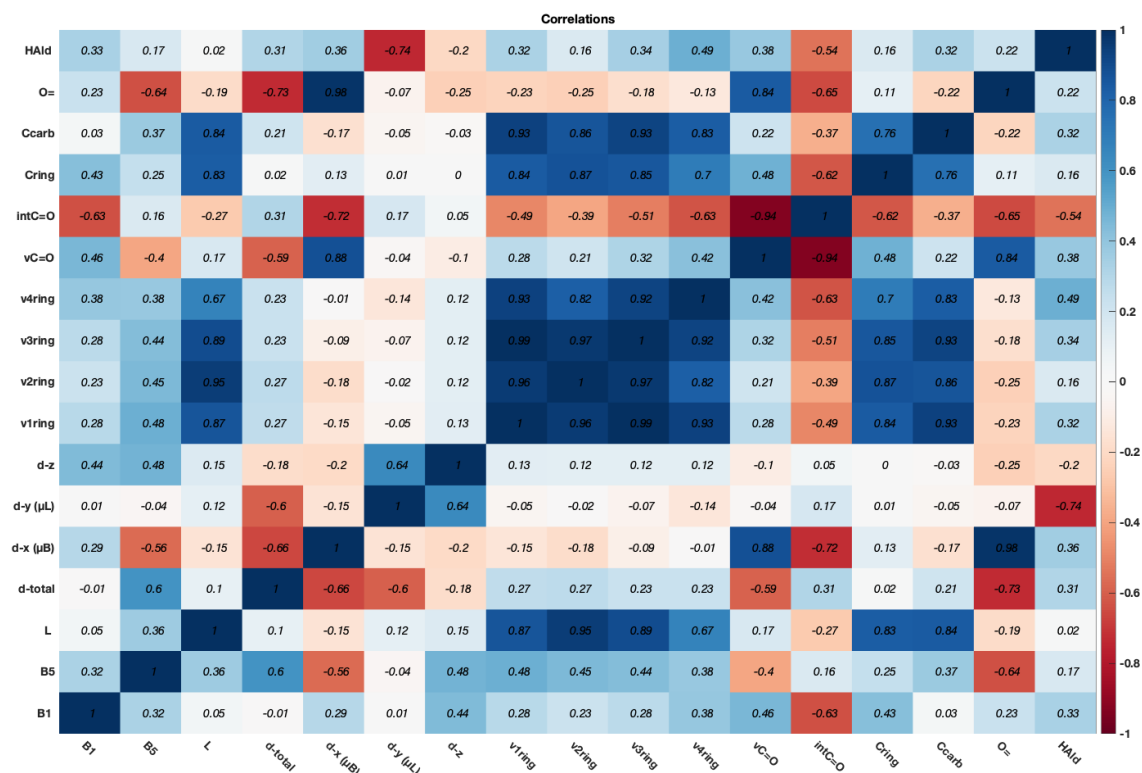
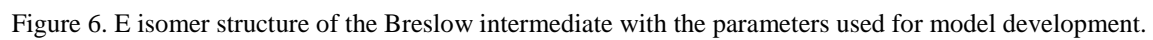
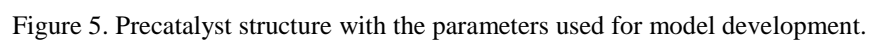


Figure 4. Correlation matrix for varying catalysts model (in figure 4 of the main text).



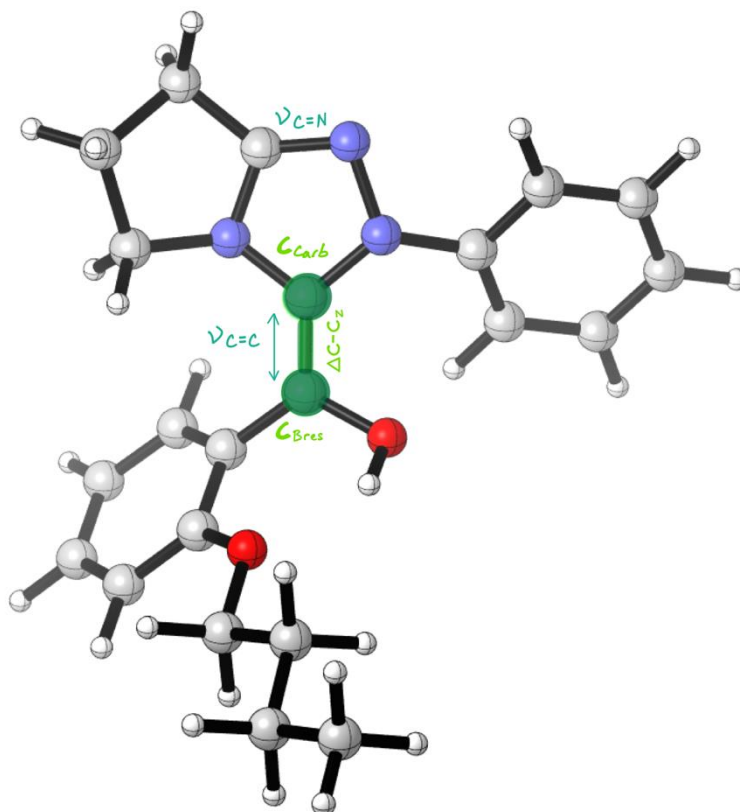
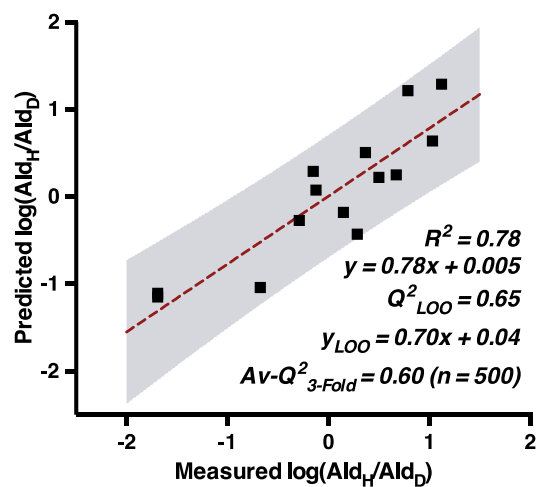


Figure 7. Z isomer structure of the Breslow intermediate with the parameters used for model development.



$$\log(\text{Ald}_H/\text{Ald}_D) = 0.02 + 0.27\Delta C-C_{\text{BresZ}} - 0.78\Delta C-H_{\text{PreCat}}$$

Figure 8. Model for varying catalysts using the Z isomer of the Breslow intermediate (in figure 4 of the main text).

## Multinomial Logistic Regression

The multinomial logistic regression models were calculated using R (V 3.6.0)<sup>18</sup> and RStudio<sup>19</sup> and the package nnet<sup>20</sup>, graphic representations were produced using the package ggplot2<sup>21</sup>. Other packages used in code are: tidyr<sup>22</sup>, reshape2<sup>23</sup>, scales<sup>24</sup>, tibble<sup>25</sup>, caret<sup>26</sup>, plyr<sup>27</sup>, dplyr<sup>28</sup> and data.table<sup>29</sup>. Variable selection was first based on a stepwise selection process (inner function of R) by applying different statistical criteria upon a scaled set of parameters which was followed by a manual inspection and estimation of success based on calculated accuracy, variable relative importance<sup>26</sup>, McFadden's pseudo R squared<sup>30</sup> and model coefficients, with a primary aim at achieving the best performance with the smallest number of possible variables to avoid over fitting. All models were calculated using the nnet::multinom function for a maximum of 2000 iterations. Each model was 3 folds cross-validated over 500 iterations, returning an average accuracy, best/worst accuracies and best/worst classification tables. The cross-validation is performed with the division of the complete set into the chosen number of folds followed by training of a model based on all folds but one and testing upon the remaining fold. A complete probabilities table for this validation is constructed by concatenating the predicted probabilities for each test set and its accuracy is computed by the division of all true (correct cases) by the total number of observations and a classification table is produced by comparing predicted classes with experimental results. Multinomial logistic regression is a generalization of the binary logistic regression (treating categorical classes as a set of independent binary regressions). In its most basic construct, logistic regression aims at performing binary classifications based on a setup that assumes a linear relationship between the Log(odds) (from which a probability is deducible) of a certain case to be in each class and a set of variables (predictors,  $X_i$  in eq. 1). Under this, logistic models present an equation with variables and corresponding coefficients ( $\beta_i$  in eq. 1). By the logistic nature of such equations, coefficients reflect the direct effect of each variable on the outcome, and therefore, on the probability (eq. 2) of a given case to be in each of the classes.

$$(1) \quad \log_b \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$



$$(2) \quad \left( \frac{p}{1-p} \right) = b^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}$$

The scaling of a data set (normalization of parameters) allows the comparison of variables by their resulting coefficients, which in turn may be used as a general tool in distinguishing strong and weak predictors in the parameter selection process. The variable importance tool employs this reasoning with the summation of the absolute value of each variable's coefficients as a relative measure (see results section – variable importance). To emphasize their relative weight in each classification, model coefficients were presented in both their log value and after applying an exponential.

A classification table and a probabilities heat map were produced for each model. The classification table includes the following information: Accuracy (blue cell) – defined as the number of correct classifications divided by the total number of observations. Computationally, the sum over the diagonal divided by the total. Precision (yellow cells) – defined as the number of correctly classified predictions under a certain class divided by the total number of predictions classified as the same class (true positive by true positive + false positive). Recall (prediction cells) – defined as the number of correctly classified predictions under a certain class divided by the total number of experimentally observed cases in the same class (true positive by true positive + false negative). As there are three classes, the complimentary fractions of the recall attribute (false negative by true positive + false negative) are meaningful, as they allow the understanding of model misclassifications in terms of which classes are wrongfully predicted.

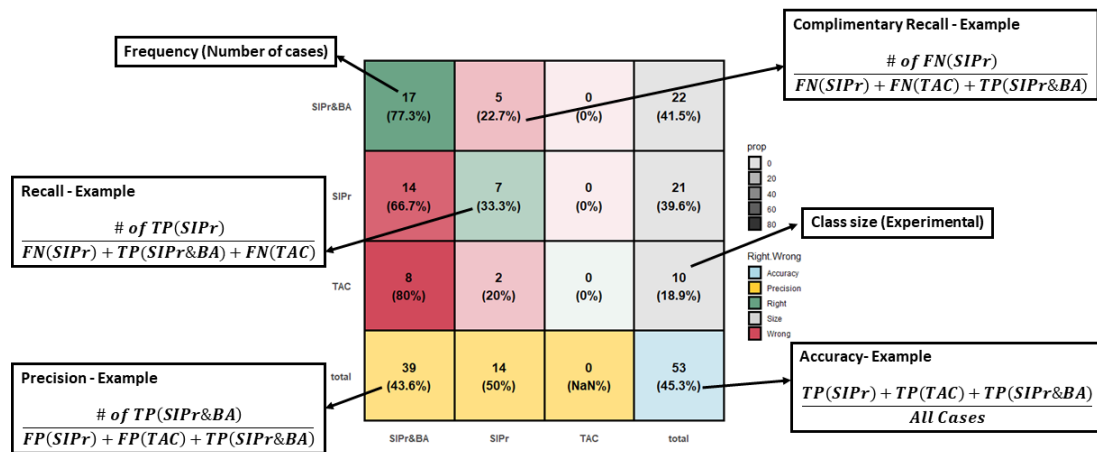
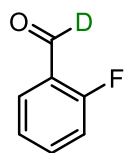


Figure 9. Graphic representations of different classification table attributes. TP = True Positive, FN = False Negative & FP = False Positive.

The probabilities heat map represents the predicted probability of each observation with a scale of 0-100% colored white to blue respectively. The experimental outcome is mentioned in a specific column and each group is color coded for convenience. Substrate names are colored green for correct classifications, orange for incorrect classifications where the true class is given the second probability out of the three possible classes, and red for a classification where the true class is given the lowest probability. A variety of models were constructed using this procedure, examining different aspects of the system. The NPA charges of the carbonyl carbon and oxygen, and of the carbon on the ring that is attached to the carbonyl were calculated. The latter two were not used as standalone variables in the following parameter selection processes because they were not found to provide predictive models. However, it was found that the charge difference of the carbonyl's carbon and oxygen as a new parameter ( $\Delta C = O$ ) is actually a better predictor than any single NPA charge. The results for the published model are presented in the following section. All of the optimized structures used in parameter estimation can be found (in .xyz format), along with the full parameter data set, source code and an additional text with all of the parameter selection process results on our GitHub page. CYLview<sup>31</sup> and IQmol<sup>32</sup> were used in molecular graphic representations.

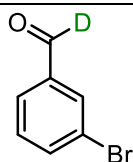
## Substrate Key for Multinomial Logistic Regression

### Substrates Classified Under Class 1 – SIPr&BA



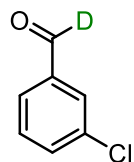
2-Fluorobenzaldehyde-*d*

2-F



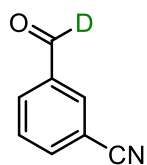
3-Bromobenzaldehyde-*d*

3-Br



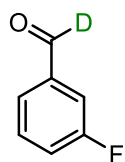
3-Chlorobenzaldehyde-*d*

3-Cl



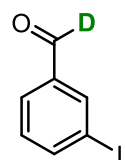
3-Cyanobenzaldehyde-*d*

3-CN



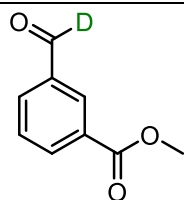
3-Fluorobenzaldehyde-*d*

3-F



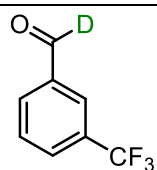
3-Iodobenzaldehyde-*d*

3-I



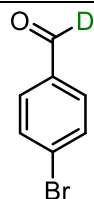
Methyl 3-formylbenzoate-*d*

3-CO<sub>2</sub>Me



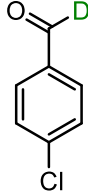
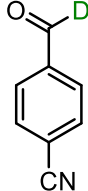
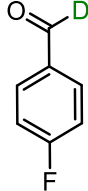
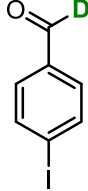
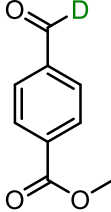
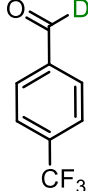
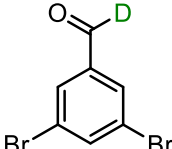
3-(Trifluoromethyl)benzaldehyde-*d*

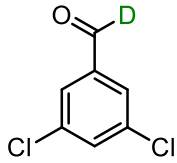
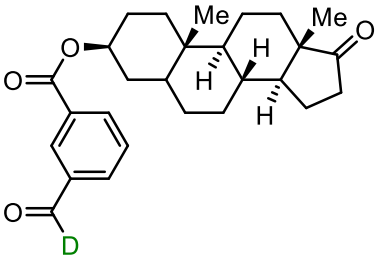
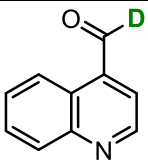
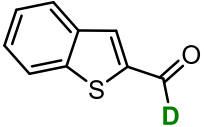
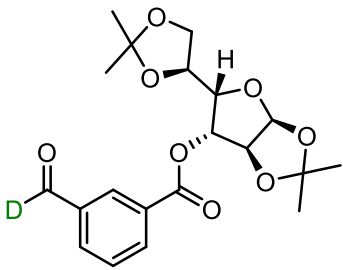
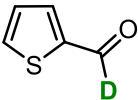
3-CF<sub>3</sub>



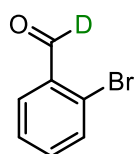
4-Bromobenzaldehyde-*d*

4-Br

	4-Chlorobenzaldehyde- <i>d</i>	4-Cl
	4-Cyanobenzaldehyde- <i>d</i>	4-CN
	4-Fluorobenzaldehyde- <i>d</i>	4-F
	4-Iodobenzaldehyde- <i>d</i>	4-I
	Methyl 4-formylbenzoate- <i>d</i>	4-CO <sub>2</sub> Me
	4-(Trifluoromethyl)benzaldehyde- <i>d</i>	4-CF <sub>3</sub>
	3,5-Dibromobenzaldehyde- <i>d</i>	3,5-Br

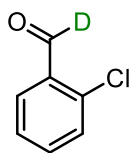
	3,5-Dichlorobenzaldehyde- <i>d</i>	3,5-Cl
	Epiandrosterone Derivative- <i>d</i>	6a
	Quinoline-4-carbaldehyde- <i>d</i>	6b
	Benzo[b]thiophene-2-carbaldehyde- <i>d</i>	6c
	Diacetone-glucose Derivative- <i>d</i>	6d
	Thiophene-2-carbaldehyde- <i>d</i>	6e

### Substrates Classified Under Class 2 – SIPr



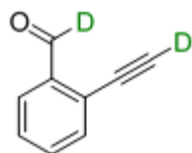
2-Bromobenzaldehyde-*d*

2-Br



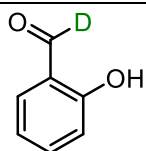
2-Chlorobenzaldehyde-*d*

2-Cl



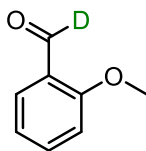
2-Ethynylbenzaldehyde-*d*

2-C≡CH



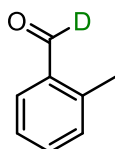
2-Hydroxybenzaldehyde-*d*

2-OH



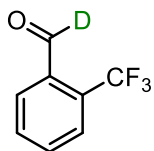
2-Methoxybenzaldehyde-*d*

2-OMe



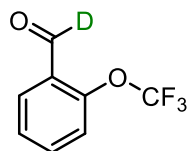
2-Methylbenzaldehyde-*d*

2-Me



2-(Trifluoromethyl)-  
benzaldehyde-*d*

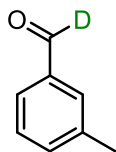
2-CF<sub>3</sub>



2-(Trifluoromethoxy)-

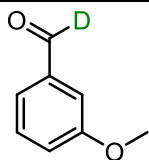
2-OCF<sub>3</sub>

benzaldehyde-*d*



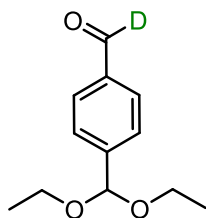
3-Methylbenzaldehyde-*d*

3-Me



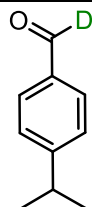
3-Methoxybenzaldehyde-*d*

3-OMe



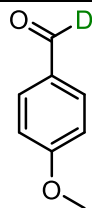
4-(Diethoxymethyl)-  
benzaldehyde-*d*

4-CH(OEt)<sub>2</sub>



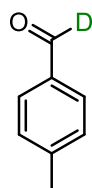
4-Isopropylbenzaldehyde-*d*

4-iPr



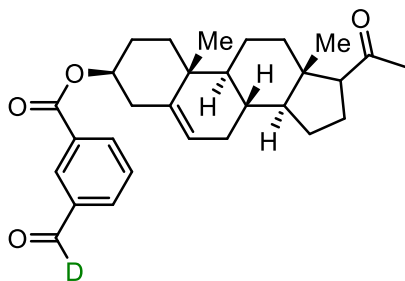
4-Methoxybenzaldehyde-*d*

4-OMe



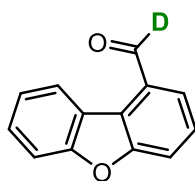
4-Methylbenzaldehyde-*d*

4-Me



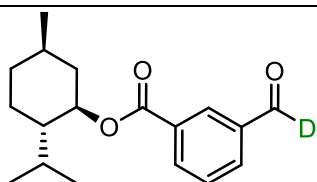
Pregnenolone Derivative-*d*

6f



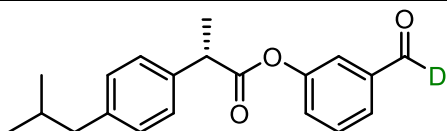
Dibenzo[b,d]furan-1-carbaldehyde-*d*

6g



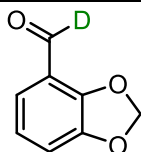
Menthol Derivative-*d*

6h



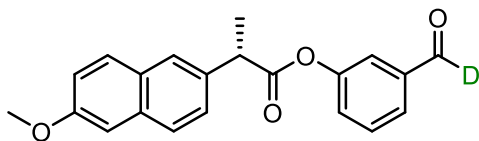
Ibuprofen Derivative-*d*

6i



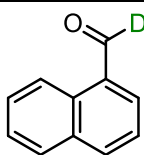
Benzo[d][1,3]dioxole-4-carbaldehyde-*d*

6j



Naproxen Derivative-*d*

6k

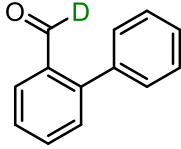
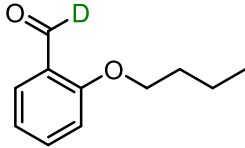
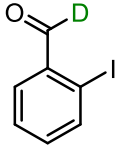
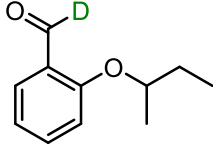
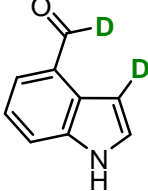
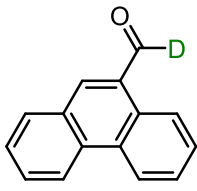
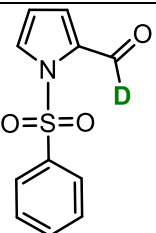


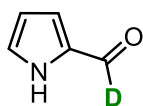
1-Naphthaldehyde-*d*

6l



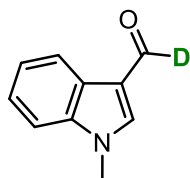
### Substrates Classified Under Class 3 – TAC

	[1,1'-biphenyl]-2-carbaldehyde- <i>d</i>	6m
	2- <i>n</i> Butoxybenzaldehyde- <i>d</i>	6n
	2-Iodobenzaldehyde- <i>d</i>	6o
	2-( <i>sec</i> Butoxy)benzaldehyde- <i>d</i>	6p
	Indole-4-carbaldehyde- <i>d</i>	6q
	Phenanthrene-9-carbaldehyde- <i>d</i>	6r
	1-(phenylsulfonyl)-pyrrole-2-carbaldehyde- <i>d</i>	6t



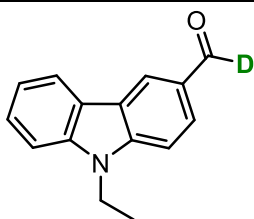
Pyrrole-2-carbaldehyde-*d*

6u



1-Methyl-indole-3-carbaldehyde-*d*

6v



9-Ethyl-carbazole-3-carbaldehyde-*d*

6s

## Published Model - Results

**formula = class ~ C=O + L + B5 + diff\_Ca\_O**

**K-Fold CV - 3 folds for 1000 iterations**

over.all.accuracy	best	worst
71.37	86.79	50.94

Best (left) and Worst (right) Classification Tables

actual	SIPr&BA	SIPr	TAC		actual	SIPr&BA	SIPr	TAC
SIPr&BA	18	2	2	***	SIPr&BA	19	2	1
SIPr	2	13	6	***	SIPr	2	15	4
TAC	0	4	6	***	TAC	0	6	4

**Entire set model - accuracy, classification table and probabilities heat map**

Accuracy	McFadden_R2
90.57%	0.644

Variable Importance

	Overall
C.O	12.11687
L	8.75129
B5	14.71319
diff_Ca_O	11.09544

Coefficients

	(Intercept)	C.O	L	B5	diff_Ca_O
2	3.2379893	-4.543052	-4.379131	6.776789	5.291028
3	0.2590097	-7.573816	-4.372160	7.936401	5.804408

exp(coefficients)

	(Intercept)	C.O	L	B5	diff_Ca_O
2	25.482433	0.0106409	0.0125363	877.2476	198.5473
3	1.295646	0.0005137	0.0126239	2797.2759	331.7588

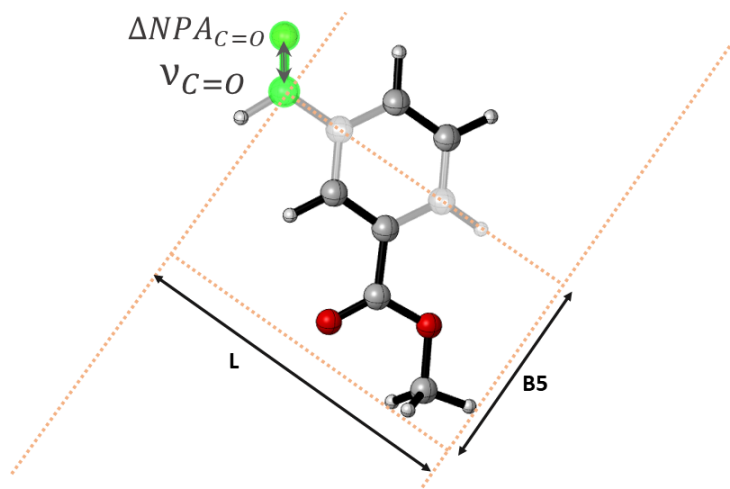


Figure 10. Representation of multinomial logistic regression parameters – C=O stretch vibration, carbonyl's carbon and oxygen NPA charge difference ( $\Delta NPA_{C=O}$ ) and Sterimol's  $L$  and  $B5$ .

SIPr&BA	22 (100%)	0 (0%)	0 (0%)	22 (41.5%)
SIPr	1 (4.8%)	19 (90.5%)	1 (4.8%)	21 (39.6%)
TAC	0 (0%)	3 (30%)	7 (70%)	10 (18.9%)
total	23 (95.7%)	22 (86.4%)	8 (87.5%)	53 (90.6%)
	SIPr&BA	SIPr	TAC	total

prop

0  
25  
50  
75  
100

Right.Wrong

Accuracy  
Precision  
Right  
Size  
Wrong

Figure 11. Classification table of the published model as it appears in R output.

	SIPr&BA	SIPr	TAC	Exp
2-F	53.58	44.86	1.56	SIPr&BA
3-Br	82.03	17.94	0.04	SIPr&BA
3-Cl	89.04	10.94	0.02	SIPr&BA
3-CN	99.47	0.53	0	SIPr&BA
3-F	99.38	0.62	0	SIPr&BA
3-I	64.57	35.33	0.1	SIPr&BA
3-CO2Me	71.48	28.34	0.18	SIPr&BA
3-CF3	94.75	5.24	0.01	SIPr&BA
4-Br	99.88	0.12	0	SIPr&BA
4-Cl	99.5	0.5	0	SIPr&BA
4-CN	100	0	0	SIPr&BA
4-F	60.34	39.42	0.24	SIPr&BA
4-I	99.95	0.05	0	SIPr&BA
4-CO2Me	100	0	0	SIPr&BA
4-CF3	100	0	0	SIPr&BA
3,5-Br	99.88	0.12	0	SIPr&BA
3,5-Cl	99.93	0.07	0	SIPr&BA
6a	70.31	27.55	2.14	SIPr&BA
6b	98.88	1.11	0.01	SIPr&BA
6c	100	0	0	SIPr&BA
6d	86.19	13.56	0.25	SIPr&BA
6e	55.92	32.2	11.88	SIPr&BA
2-Br	8.09	86.68	5.22	SIPr
2-Cl	13.29	81.75	4.96	SIPr
2-C=CH	1.99	92.94	5.07	SIPr
2-OH	0	63.65	36.35	SIPr
2-Ome	0	29.38	70.62	SIPr
2-Me	0	58.75	41.25	SIPr
2-CF3	98.15	1.83	0.02	SIPr
2-OCF3	0.48	89.19	10.33	SIPr
3-Me	0.75	98.13	1.12	SIPr
3-Ome	16.9	82.6	0.5	SIPr
4-CH(OEt)2	28.85	70.03	1.12	SIPr
4-iPr	48.27	50.61	1.12	SIPr
4-Ome	0.31	89	10.69	SIPr
4-Me	11.73	86.86	1.41	SIPr
6f	0.01	88.94	11.04	SIPr
6g	0	76.01	23.99	SIPr
6h	43.33	55.54	1.13	SIPr
6i	0	90.9	9.1	SIPr
6j	0.18	89.75	10.07	SIPr
6k	0.01	86.15	13.84	SIPr
6l	0	50.47	49.53	SIPr
6m	0	49.87	50.13	TAC
6n	0	4.41	95.59	TAC
6o	2.5	91.33	6.16	TAC
6p	0	10.8	89.2	TAC
6q	0	67.5	32.5	TAC
6r	0.02	54.78	45.2	TAC
6s	0	36.19	63.81	TAC
6t	0.02	6.99	92.99	TAC
6u	0	0.13	99.87	TAC
6v	0	0.37	99.63	TAC

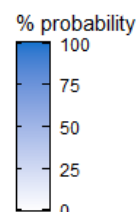


Figure 12. Probabilities heat map table of the published model as it appears in R output.

## DFT Calculations

### Computational method

All gas phase geometry optimizations were carried out at the M06-2X<sup>2</sup>/6-31G(d,p) level using Gaussian16 software package.<sup>1</sup> This combination has been shown to be appropriate for mechanistic computations including hydrogen bonds and reaction barriers.<sup>33-36</sup> Frequency analyses were performed to verify the energy minima and transition states, and intrinsic reaction coordinate (IRC) computations were used to confirm the elementary steps. Single point energies were computed in toluene using the SMD solvent model<sup>37</sup> at the M06-2X/6-311++G(d,p). Thermal corrections were obtained with the geometry optimization method at the experimental temperature of 295.15 K, with a pressure of 24.2 atm to match a 1M standard concentration. All the values of the article are therefore M06-2X/6-311++G(d,p)/SMD//M06-2X/6-31G(d,p) Gibbs energies. CYLview software<sup>31</sup> is employed for rendering the 3D structures of the molecules. To estimate the apparent activation barrier of the full catalytic cycle we used the energy span ( $\delta E$ ) concept.<sup>38-40</sup> The model uses the TOF determining transition state (**TDTS**) and TOF determining intermediate (**TDI**), as given in eq. 3:

$$(3) \quad \text{TOF} \approx \frac{k_B T}{h} e^{-\delta E / RT}$$

$$\delta E = \begin{cases} G_{\text{TDTS}} - G_{\text{TDI}} & \text{if the TDTS comes after the TDI} \\ G_{\text{TDTS}} - G_{\text{TDI}} + \Delta G_{\text{rxn}} & \text{if the TDTS comes before the TDI} \end{cases}$$

### Mechanism and energy profile

In the experiment, SIPr catalyst was used for the NHC-catalyzed hydrogen-deuterium exchange (**HDE**) process (Fig. 2 in the main text). For simplicity we substituted it with dimethylphenyl groups (**A<sub>c</sub>**, see Fig. 13). A very similar catalyst, **SIMes** (See Fig. 6a-b in the main text), which resembles with model in terms of steric and electronics, was also found experimentally to observe the same HDE process. We choose 2-chlorobenzaldehyde (**R**) as a model substrate due to its excellent performance in deuterium incorporation in addition to its high reaction yield.

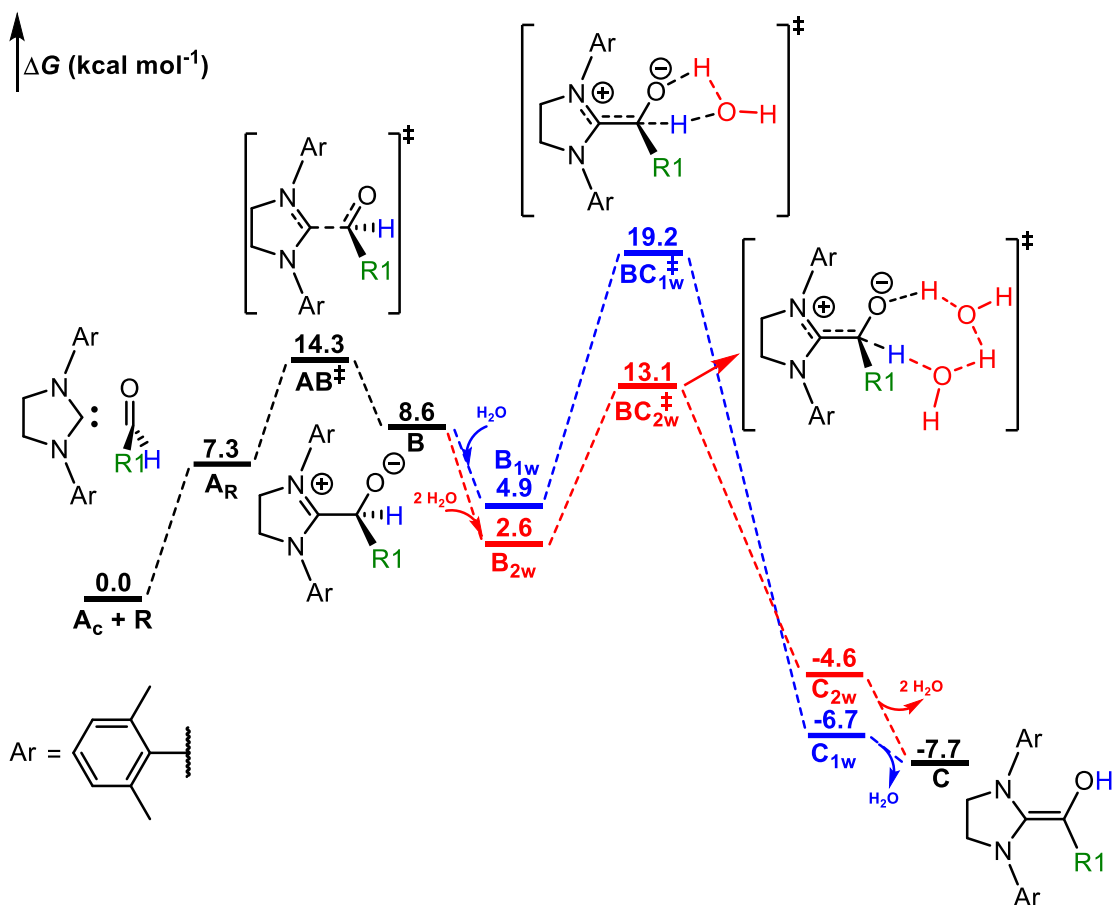


Figure 13. Gibbs free energy profile of the NHC-catalyzed C-H bond activation of 2-chlorobenzaldehyde (**R**) in the presence of water (one molecule in blue, two in red), leading to a Breslow intermediate.

Figure 13 shows the computed energy profile for the H<sub>2</sub>O-aided H-transfer through the formation of a Breslow intermediate. At first, the **A<sub>c</sub>** free carbene forms a weak van der Waals complex with **R** (**A<sub>R</sub>**, unstable when including entropic corrections), which undergoes a nucleophilic attack on the electron deficient carbonyl carbon via transition state **AB<sup>‡</sup>**, forming an acyl anion (**B**). After the formation of the H-bonded complexes with one or two H<sub>2</sub>O (**B<sub>1w</sub>** or **B<sub>2w</sub>**), the subsequent step involves a water molecule assisted Grotthuss type H-transfer<sup>43,44</sup> from the carbon to the oxygen via a concerted 5- or 7-membered ring transition state (**BC<sub>1w</sub><sup>‡</sup>** and **BC<sub>2w</sub><sup>‡</sup>**), forming first the acyl anion adduct (**C<sub>1w</sub>** and **C<sub>2w</sub>**) and finally the entropically favored Breslow intermediate (**C**). The “dry” waterless H-transfer has a prohibitive barrier (55.4 kcal mol<sup>-1</sup>, not shown in Fig. 13) and obviously cannot lead to the deuteration. Since according to our computations, the addition of more water molecules was not favorable, we concluded that the two-water pathway is

the most probable one (a three-water, nine-membered ring does not provide a suitable transition state, and a seven membered ring with another H<sub>2</sub>O connected by an H-bond is not entropically favored). Noteworthy, the involvement of H-bonded H<sub>2</sub>O units does not stabilize the Gibbs energies of **A<sub>w/2w</sub>**, **AB<sub>w/2w</sub><sup>‡</sup>** and **B<sub>w/2w</sub>** (see table 1), and therefore we only depict the waterless compounds for these.

The complete reaction involves going back and forth through the energy profile of Fig. 13, with the back reaction giving the deuterated product. In this graph only, the H pathway is computed and not the D one. The internal energy profile will look exactly the same, but clearly the inclusion of heavy water in the process will make small changes in the Gibbs energy values due to entropic and zero-point energy corrections. For the complete reaction the TDI is **C**, the Breslow intermediate, while **AB<sup>‡</sup>** results the TDTS, with an estimated  $\delta E = 22.0 \text{ kcal mol}^{-1}$ , corresponding to a TOF  $\sim 1 \text{ h}^{-1}$ . Interestingly, with only one water bridge the TDTS would be the H-transfer (**BC<sub>1w</sub><sup>‡</sup>**), with a  $\delta E = 26.9 \text{ kcal mol}^{-1}$ , and a much slower TOF  $\sim 10^{-4} \text{ h}^{-1}$ .

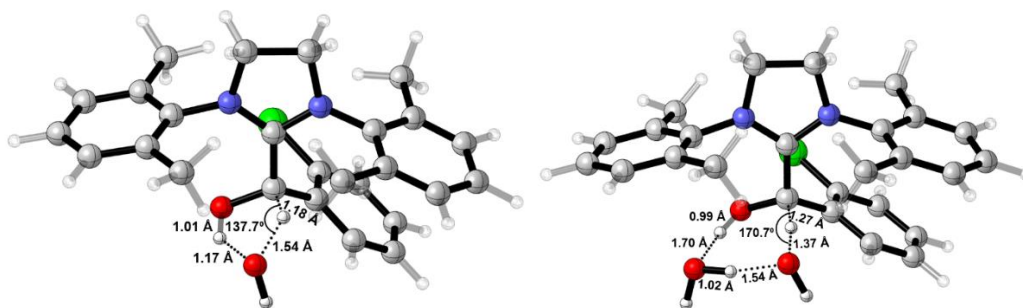


Figure 14. Geometries of the Grotthuss type H-transfer transition states (**BC<sub>1w</sub><sup>‡</sup>** left, **BC<sub>2w</sub><sup>‡</sup>** right) with selected distances and angles. The higher angles of the seven-membered ring is one of the reasons for the lower transition state with two waters.



## Alcohols as H/D source

As discussed in the main article, the studied NHC-catalyzed **HDE** reaction was also observed to be driven in the presence of relatively diluted alcohols. It can be seen from Fig. 6b in the main text that the deuterium percentage loss trend in the studied alcohols follows the following order: methanol  $\approx$  ethanol > trifluoroethanol > 4-methylphenol > phenol > 4-fluorophenol > pentafluorophenol, where the latter shows no D loss at all (i.e. no reaction). To rationalize this sharp difference, we first explored the NHC-catalyzed C-H bond activation of **R** via a simple methanol-assisted H-transfer (Fig. 15). The energy profile is very similar to the water case, with the two-MeOH pathway as the most probable one, the same TDI and TDTS, and therefore the same  $\delta E$  and TOF. Similar trend was also obtained for ethanol and trifluoroethanol (Fig. 16 and 17, respectively). Although in the latter case the reaction is feasible both with one- and two molecules of trifluoroethanol, the two-trifluoroethanol assisted H-transfer will be the thermodynamically preferred path as it goes via a more stable “protonated acyl anion adduct”<sup>41,42</sup> stabilized by alkoxide instead of the acyl anion mentioned above. The presence of more electron-withdrawing groups in the alcohol stabilizes this intermediate in conjunction with an alkoxide, instead of keeping the negative charge in the activated aldehyde. These observations lead us to study experimentally the conditions for the catalysis with alcohols, as explained in the main text.

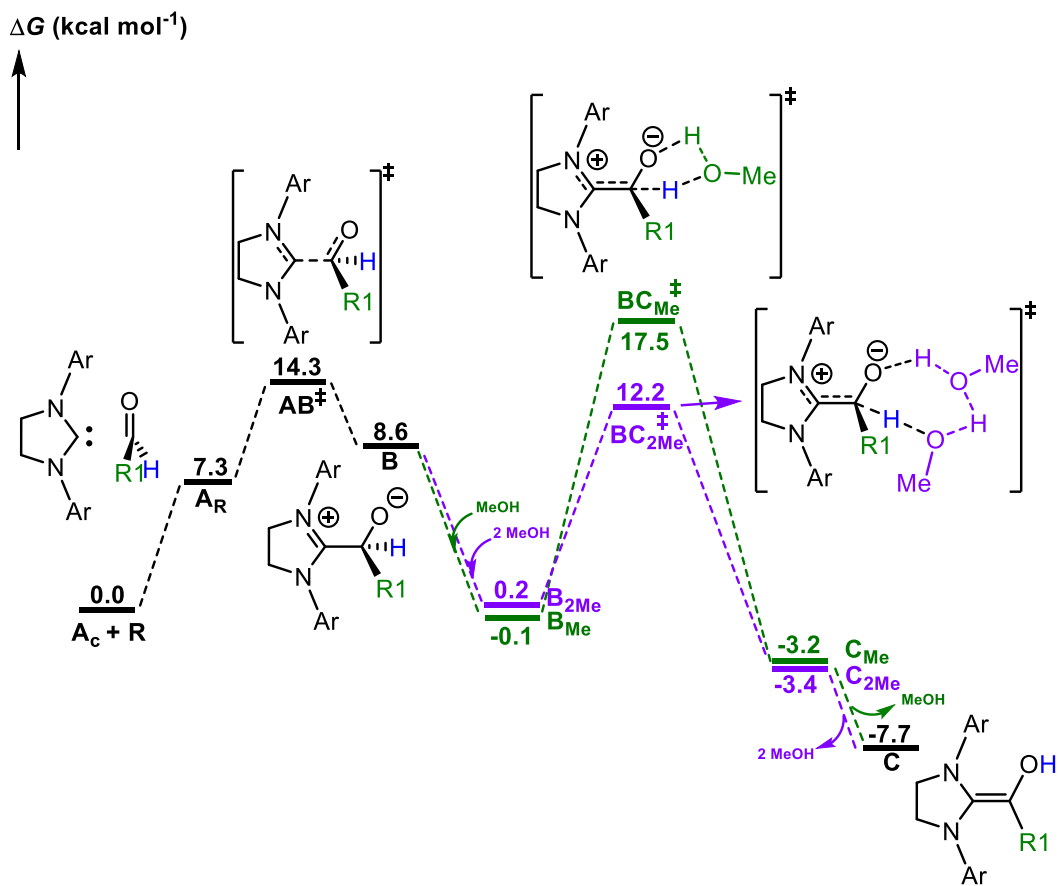


Figure 15. Gibbs free energy profile of the NHC-catalyzed C-H bond activation of 2-chlorobenzaldehyde (**R**) in the presence of one or two MeOH (green and purple, respectively), leading to a Breslow intermediate.

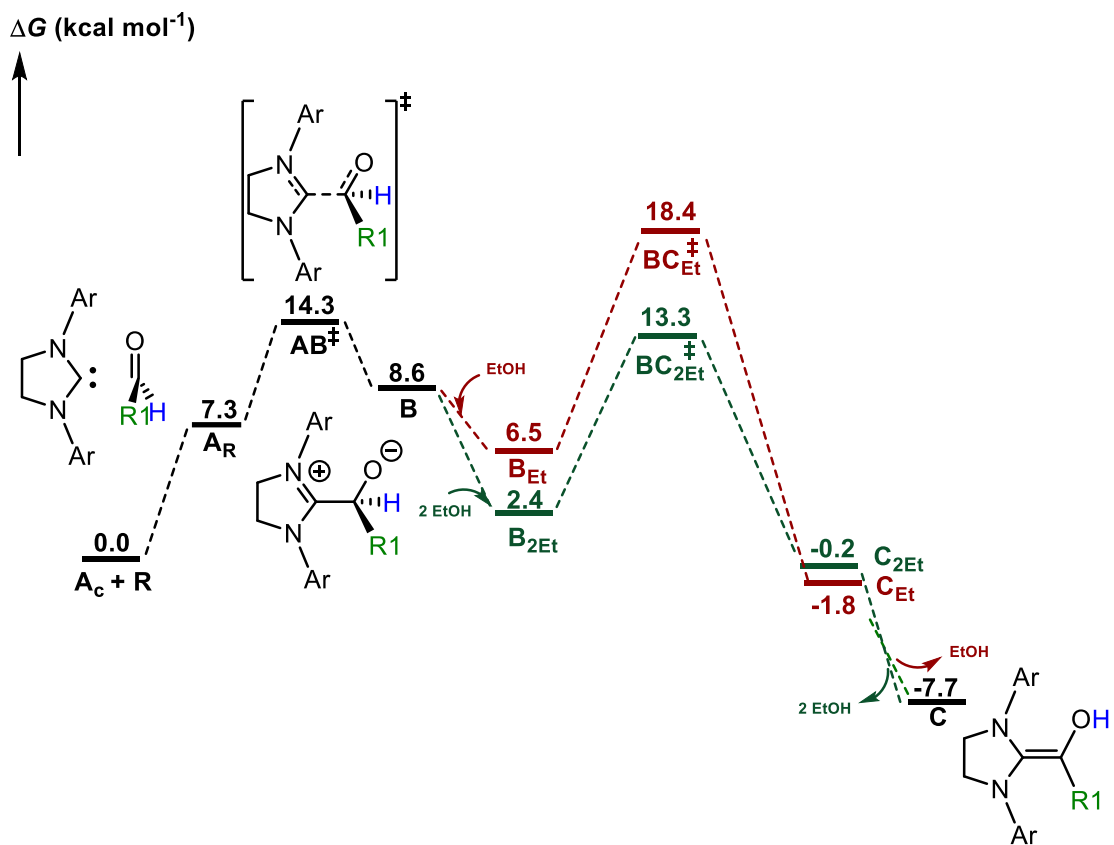


Figure 16. Gibbs free energy profile of the NHC-catalyzed C-H bond activation of 2-chlorobenzaldehyde (**R**) in the presence of one or two EtOH (brown and green, respectively), leading to a Breslow intermediate.

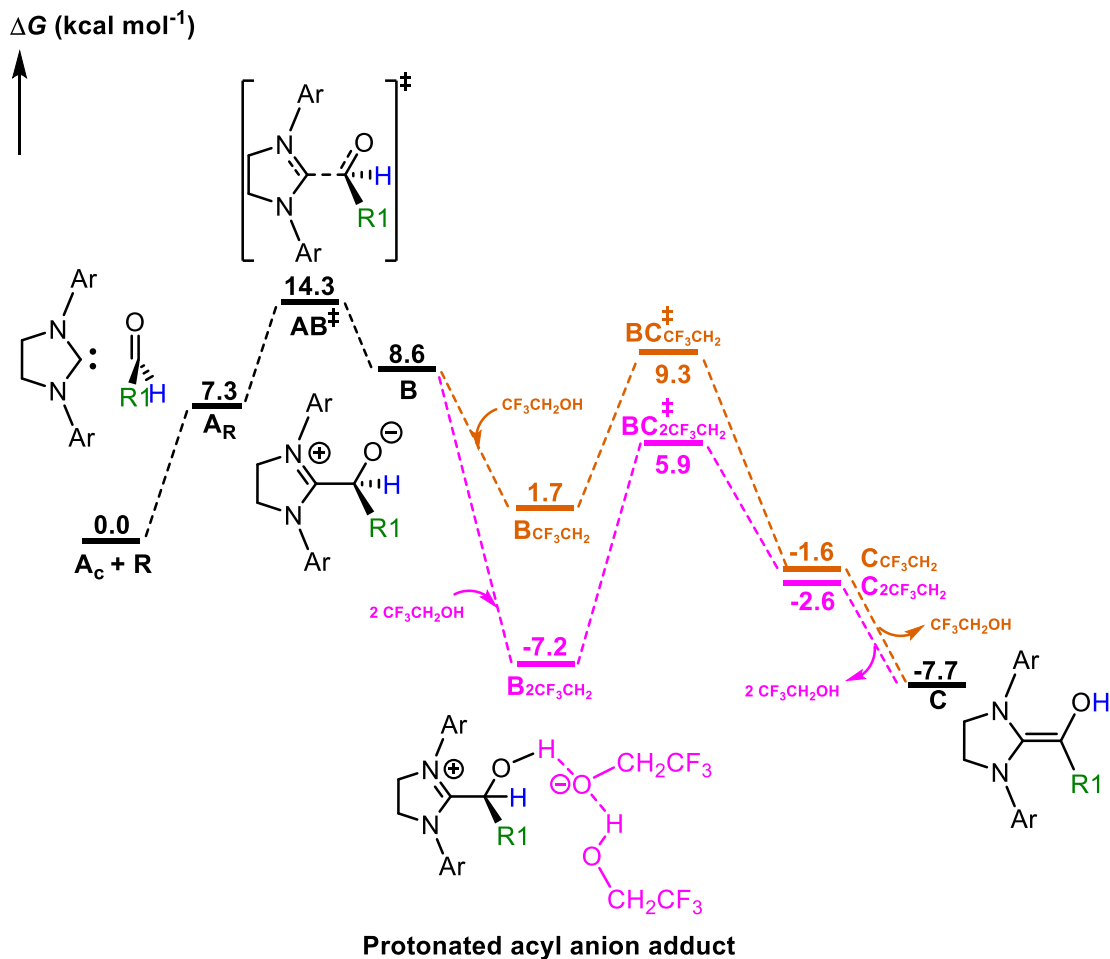


Figure 17. Gibbs free energy profile of the NHC-catalyzed C-H bond activation of 2-chlorobenzaldehyde (**R**) in the presence of one or two  $\text{CF}_3\text{CH}_2\text{OH}$  (light brown and pink, respectively), leading to the generation of a Breslow intermediate.

A similar series of computations were performed for different phenols (phenol, *p*-methylphenol, *p*-fluorophenol and pentafluorophenol) to understand the differences in their reactivities. The one-bridging phenol pathway (Fig. 18) already show a lower  $\Delta G^\ddagger$  value on the H-transfer transition state than that of MeOH, indicating that the Grotthuss mechanism is actually enhanced with lower  $\text{pK}_a$  phenols. The reactant and product states of this step with phenols (**B<sub>x</sub>** and **C<sub>x</sub>**) form stabler hydrogen bonded adducts than water and methanol (see Figs. 18, 13 and 15). While for  $\text{H}_2\text{O}$ , methanol, ethanol and trifluoroethanol the TDI is **C**, for most of the phenols it is **C<sub>x</sub>**, where the H-bond wins the

contest against entropy. However, pentafluorophenol, due to the presence of its highly electron withdrawing groups, forms an extremely stable protonated acyl anion adduct (**B<sub>x</sub>**, see Fig. 18).

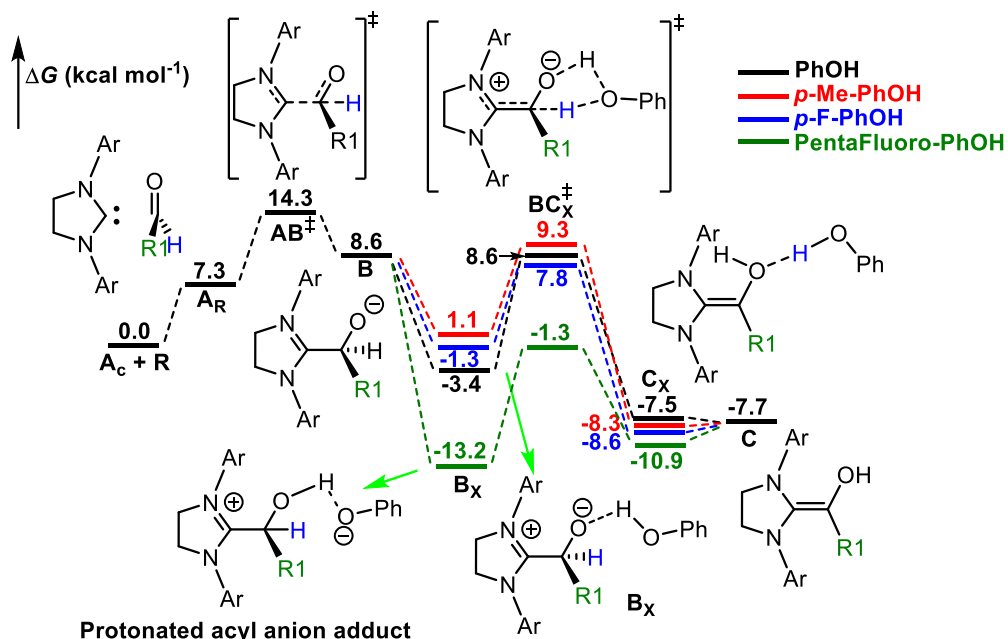


Figure 18. Gibbs free energy profile of the NHC-catalyzed C-H bond activation of 2-chlorobenzaldehyde (**R**) in the presence of one phenol (the 'X' represents the different phenols).

As a result, the  $\delta E$  is slightly higher for most of the phenols, which can explain the lower yield, along with the loss in deuterium incorporation. As shown in Fig. 6b of the main text, pentafluorophenol was completely inactive to deuteration, despite its low  $pK_a$  which would, naively, suggest that the H/D substitution might be enhanced. But the TDI (the protonated acyl anion complex) is low enough to virtually stop the deuteration. This would provide a further proof of this adduct stability effect. However, the two-phenol and its substituted phenol-assisted (**B<sub>2x</sub>**) H-transfer reaction are again favorable compared to the one-phenol mechanism (Fig. 19). In all these cases the TDI results the low-lying protonated acyl anion adduct (**B<sub>2x</sub>**), and added to the lower energies for the H-transfer transition state, this leads to the two-phenol mechanism as the most probable one. As can be seen from Fig.

19, there is a correlation between the stability of **B<sub>2</sub>x** and the pK<sub>a</sub> of the phenol, which broadly correlates with the experimental activity (see Fig. 6b of the main text). Even if the H-transfer indeed is facilitated by the low pK<sub>a</sub>, this is irrelevant as the aldehyde activation (**AB<sup>‡</sup>**) is still the TDTS, and this transition state is not dependent on the hydrogen source. Due to the extremely low TDI of pentafluorophenol (the “thermodynamic sink”, as termed in the main text), its resulting  $\delta E$  is prohibitive (36.5 kcal mol<sup>-1</sup>), as observed experimentally. On the other hand, the  $\delta E$  obtained with *p*-Me-PhOH (23.4 kcal mol<sup>-1</sup>, TOF ~ 0.1 h<sup>-1</sup>) indicates a slow but favorable reaction, supporting the experimental result (Fig. 6b of the main text). For the PhOH and *p*-F-PhOH cases the energy spans are 24.6 and 26.5 kcal mol<sup>-1</sup>, respectively, which would lead to too slow reactions at the experimental temperatures (TOFs of the order of 10<sup>-2</sup> and 10<sup>-4</sup> h<sup>-1</sup>). However, the experiments show a 11% and 6% deuterium loss with these phenols. Noteworthy, the DFT and especially the solvation entropy errors are of the same magnitude to the experimental/computational disparity, and therefore the confidence interval of the computations may well lie within the experimental ones. In any case, the stability of the TDI (plus the solubility and H-bonded aggregation issue<sup>45,46</sup> as described in the main text) explains the observed deuteration trend: with phenols the protonated acyl anion adduct determines the reaction rate, instead of the Breslow intermediate as with water, methanol, ethanol and trifluoroethanol.

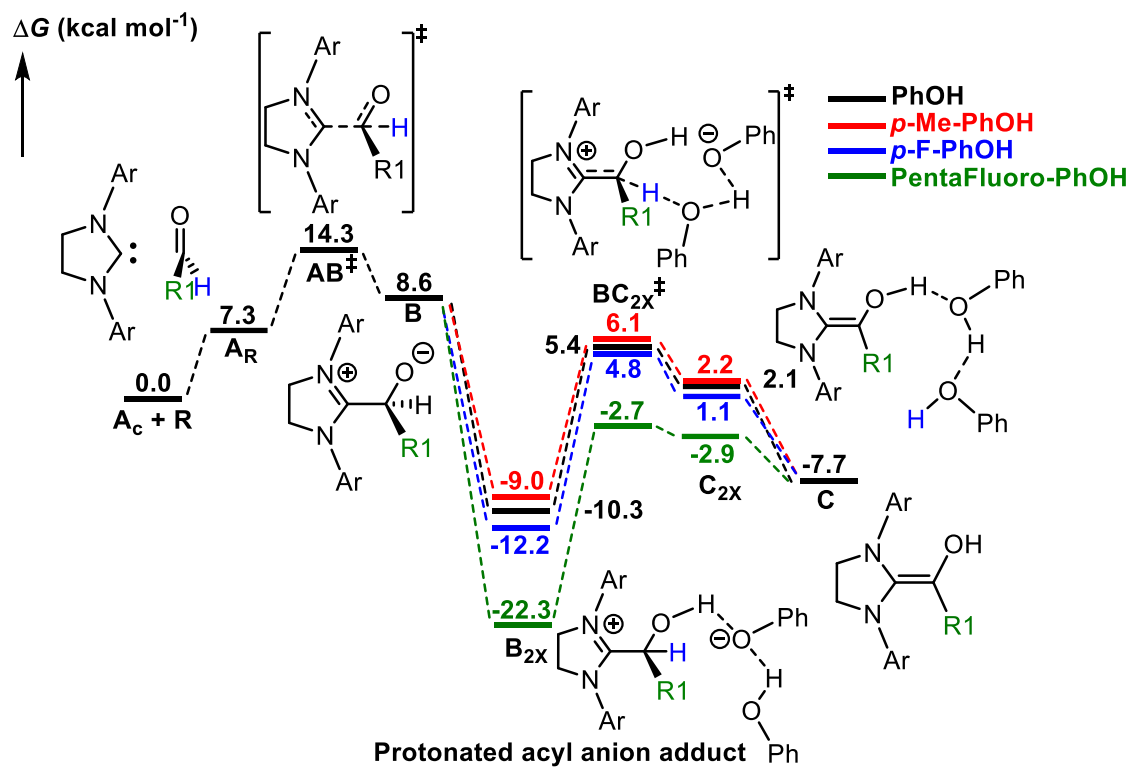


Figure 19. Gibbs free energy profile of the NHC-catalyzed C-H bond activation of 2-chlorobenzaldehyde (**R**) in the presence of two phenol and its substituted phenol molecules.

**Table 1: Energies of all the species involved in the NHC-catalyzed HDE process.**

System	E [Ha]	Thermal correction [Ha]	G [Ha]	$\Delta G$ (kcal/mol)
NHC (A <sub>c</sub> )	-846.614260	0.324255	-846.290005	
R	-805.1245153	0.072335	-805.052180	
H <sub>2</sub> O	-76.42539489	0.007261	-76.418134	
CH <sub>3</sub> OH	-115.7073411	0.032659	-115.674682	
C	-1651.775638	0.421149	-1651.35449	
CH <sub>3</sub> CH <sub>2</sub> OH	-155.017500	0.058912	-154.958588	
CF <sub>3</sub> CH <sub>2</sub> OH	-452.763463	0.031647	-452.731816	
PhOH	-307.4331106	0.080254	-307.352857	
<i>p</i> -Me-PhOH	-346.7395984	0.104719	-346.6348794	
<i>p</i> -F-PhOH	-406.672582	0.070704	-406.601878	
PentaFluoro-PhOH	-803.5911897	0.033431	-803.557759	
DRY				
A <sub>c</sub> + R	-	-	-1651.342185	0.0
A <sub>R</sub>	-1651.745892	0.415402	-1651.330490	7.3
AB <sup>‡</sup>	-1651.739099	0.419645	-1651.319454	14.3
B	-1651.747734	0.419313	-1651.328421	8.6
BC <sup>‡</sup>	-1651.682989	0.416818	-1651.266171	47.7
C	-	-	-1651.354489	-7.7
H <sub>2</sub> O				
A <sub>Rw</sub>	-1728.185935	0.438199	-1727.747736	7.9
AB <sub>w</sub> <sup>‡</sup>	-1728.180115	0.442782	-1727.737333	14.4
B <sub>w</sub>	-1728.196104	0.443545	-1727.752559	4.9
BC <sub>w</sub> <sup>‡</sup>	-1728.173353	0.443635	-1727.729718	19.2
C <sub>w</sub>	-1728.214120	0.443169	-1727.770951	-6.7
2 H <sub>2</sub> O				
A <sub>R2w</sub>	-1804.62096	0.462692	-1804.158268	12.7
AB <sub>2w</sub> <sup>‡</sup>	-1804.619518	0.464128	-1804.15539	14.5
B <sub>2w</sub>	-1804.641881	0.467643	-1804.174238	2.6
BC <sub>2w</sub> <sup>‡</sup>	-1804.622852	0.465268	-1804.157584	13.1
C <sub>2w</sub>	-1804.646983	0.461177	-1804.185806	-4.6
MeOH				
A <sub>RMe</sub>	-1767.467991	0.464786	-1767.003205	8.6
AB <sub>Me</sub> <sup>‡</sup>	-1767.466778	0.468421	-1766.998357	11.6
B <sub>Me</sub>	-1767.487288	0.470469	-1767.016819	-0.1



BC <sub>Me</sub> <sup>‡</sup>	-1767.458709	0.469698	-1766.989011	17.5
C <sub>Me</sub>	-1767.492939	0.471038	-1767.021901	-3.2
2 MeOH				
A <sub>R2Me</sub>	-1883.187293	0.514621	-1882.672672	11.8
AB <sub>2Me</sub> <sup>‡</sup>	-1883.185079	0.517942	-1882.667137	15.3
B <sub>2Me</sub>	-1883.208922	0.518397	-1882.690525	0.2
BC <sub>2Me</sub> <sup>‡</sup>	-1883.190033	0.517870	-1882.672163	12.2
C <sub>2Me</sub>	-1883.218454	0.521550	-1882.696904	-3.4
EtOH				
B <sub>Et</sub>	-1806.789907	0.499477	-1806.290430	6.5
BC <sub>Et</sub> <sup>‡</sup>	-1806.769338	0.497847	-1806.271491	18.4
C <sub>Et</sub>	-1806.801159	0.497514	-1806.303645	-1.8
2 EtOH				
B <sub>2Et</sub>	-1961.829614	0.574148	-1961.255466	2.4
BC <sub>2Et</sub> <sup>‡</sup>	-1961.810295	0.572095	-1961.238200	13.3
C <sub>2Et</sub>	-1961.836721	0.577086	-1961.259635	-0.2
CF <sub>2</sub> CH <sub>2</sub> OH				
B <sub>CF3CH2</sub>	-2104.546707	0.475418	-2104.071289	1.7
B <sub>CF3CH2</sub> <sup>‡</sup>	-2104.531135	0.471975	-2104.059160	9.3
C <sub>CF3CH2</sub>	-2104.553032	0.476515	-2104.076517	-1.6
2 CF <sub>2</sub> CH <sub>2</sub> OH				
B <sub>2CF3CH2</sub>	-2557.338914	0.521562	-2556.817352	-7.2
BC <sub>2CF3CH2</sub> <sup>‡</sup>	-2557.317156	0.520666	-2556.796490	5.9
C <sub>2CF3CH2</sub>	-2557.334468	0.524577	-2556.809891	-2.6
PhOH				
B <sub>Ph</sub>	-1959.218690	0.518260	-1958.700430	-3.4
BC <sub>Ph</sub> <sup>‡</sup>	-1959.200271	0.518986	-1958.681285	8.6
C <sub>Ph</sub>	-1959.225713	0.518705	-1958.707008	-7.5
<i>p</i> -Me-PhOH				
B <sub><i>p</i>-Me</sub>	-1998.521537	0.546226	-1997.975311	1.1
BC <sub><i>p</i>-Me</sub> <sup>‡</sup>	-1998.505613	0.543368	-1997.962245	9.3
C <sub><i>p</i>-Me</sub>	-1998.534270	0.543920	-1997.990350	-8.3
<i>p</i> -F-PhOH				
B <sub><i>p</i>-F</sub>	-2058.457503	0.511383	-2057.946120	-1.3
BC <sub><i>p</i>-F</sub> <sup>‡</sup>	-2058.441322	0.509739	-2057.931583	7.8
C <sub><i>p</i>-F</sub>	-2058.468231	0.510450	-2057.957781	-8.6

PentaFluoro-PhOH				
B <sub>PentaF</sub>	-2455.399527	0.478509	-2454.921018	-13.2
BC <sub>PentaF</sub> <sup>‡</sup>	-2455.375312	0.473292	-2454.90202	-1.3
C <sub>PentaF</sub>	-2455.392642	0.475284	-2454.917358	-10.9
2 PhOH				
B <sub>2Ph</sub>	-2266.687040	0.622666	-2266.064374	-10.3
BC <sub>2Ph</sub> <sup>‡</sup>	-2266.653361	0.614107	-2266.039254	5.4
C <sub>2Ph</sub>	-2266.662680	0.618119	-2266.044561	2.1
2 <i>p</i> -Me-PhOH				
B <sub>2<i>p</i>-Me</sub>	-2345.298601	0.672369	-2344.626232	-9.0
BC <sub>2<i>p</i>-Me</sub> <sup>‡</sup>	-2345.265690	0.663474	-2344.602216	6.1
C <sub>2<i>p</i>-Me</sub>	-2345.275511	0.666999	-2344.608512	2.2
2 <i>p</i> -F-PhOH				
B <sub>2<i>p</i>-F</sub>	-2465.169204	0.603754	-2464.565450	-12.2
BC <sub>2<i>p</i>-F</sub> <sup>‡</sup>	-2465.134631	0.596297	-2464.538334	4.8
C <sub>2<i>p</i>-F</sub>	-2465.142519	0.598333	-2464.544186	1.1
2 PentaFluoro-PhOH				
B <sub>2PentaF</sub>	-3259.023929	0.530773	-3258.493176	-22.3
BC <sub>2PentaF</sub> <sup>‡</sup>	-2465.134631	0.596297	-3258.461953	-2.7
C <sub>2PentaF</sub>	-2465.142519	0.598333	-3258.462251	-2.9

## References

1. Frisch, M. J., Trucks G.W., Schlegel H. B., Scuseria G. E., Robb M. A. & Cheeseman J. R. Gaussian 16 Rev. B.01. Wallingford, C. T. 2016.
2. Zhao, Y. & Truhlar, D.G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215-241 (2007).
3. Valero, R., Gomes J. R. B., Truhlar D. G. & Illas, F. Good performance of the M06 family of hybrid meta generalized gradient approximation density functionals on a difficult case: CO adsorption on MgO(001). *J. Chem. Phys.* **129**, 124710 (2008).
4. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297-3305 (2005).
5. Weigend, F. Accurate coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **8**, 1057-1065 (2006).
6. Merrick, J. P., Moran, D. & Radom, L. An evaluation of harmonic vibrational frequency scale factors. *J. Phys. Chem. A* **111** (45), 11683-11700 (2007).
7. Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors – the Case of Sterimol Steric Parameters. *ACS Catal.* **9**, 2313-2323 (2019).
8. Verloop, A. & Tipker, J. Use of linear free energy related and other parameters in the study of fungicidal selectivity. *Pesticide Science: Wiley Online Library* 379-390 (1976).
9. Glendening, E. D., Reed, A. E., Carpenter, J. E. & Weinhold, F. NBO Version 3.1.
10. Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S., A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis. *Science* **347** (6223), 737-43 (2015).

11. Santiago, C. B.; Milo, A.; Sigman, M. S., Developing a Modern Approach To Account for Steric Effects in Hammett-Type Correlations. *J. Am. Chem. Soc.* **138**, 13424-13430 (2016).
12. Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S., A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis. *Science* Vol. **347**, pp 737-743 (2015).
13. Milo, A.; Bess, E. N.; Sigman, M. S., Interrogating selectivity in catalysis using molecular vibrations. *Nature* **Vol. 507**, pp 210-214 (2014).
14. Dhayalan, V.; Gadekar, S. C.; Alassad, Z.; Milo, A., Unravelling mechanistic features of organocatalysis with in situ modifications at the secondary sphere. *Nat. Chem.* **11** (6), 543-551 (2019).
15. MATLAB (The MathWorks, Natick, MA).
16. Goodman, S., A Dirty Dozen: Twelve P-Value Misconceptions. In *Seminars in Hematology*; **Vol. 45**, pp 135-140 (2008).
17. Guo, J.-Y.; Minko, Y.; Santiago, C. B.; Sigman, M. S., Developing Comprehensive Computational Parameter Sets To Describe the Performance of Pyridine-Oxazoline and Related Ligands. *ACS Catal.* **7** (6), 4144-4151 (2017).
18. R Core Team (2014). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. URL <http://www.Rproject.org/>
19. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
20. Venables WN, Ripley BD (2002). Modern Applied Statistics with S, Fourth edition. Springer, New York. ISBN 0-387-95457-0
21. Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4
22. Hadley Wickham and Lionel Henry (2020). tidyr: Tidy Messy Data. R package version 1.1.0. <https://CRAN.R-project.org/package=tidyr>
23. Hadley Wickham (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, **21(12)**, 1-20. URL <http://www.jstatsoft.org/v21/i12/>

24. Hadley Wickham and Dana Seidel (2020). scales: Scale Functions for Visualization. R package version 1.1.1. <https://CRAN.R-project.org/package=scales>
25. Kirill Muller and Hadley Wickham (2020). tibble: Simple Data Frames. R package version 3.0.3. <https://CRAN.R-project.org/package=tibble>
26. Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. <https://CRAN.R-project.org/package=caret>
27. Hadley Wickham (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, **40(1)**, 1-29. URL <http://www.jstatsoft.org/v40/i01/>.
28. Hadley Wickham, Romain Francois, Lionel Henry and Kirill Muller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.0. <https://CRAN.R-project.org/package=dplyr>
29. Matt Dowle and Arun Srinivasan (2019). data.table: Extension of data.frame. R package version 1.12.8. <https://CRAN.R-project.org/package=data.table>
30. Urban Travel Demand: A Behavioral Analysis Tom Domencich and Daniel L. McFadden, North-Holland Publishing Co., 1975.
31. Legault, C. Y. CYLview, 1.0b; Universite de Sherbrooke, 2009, <http://www.cylview.org>.
32. Quantum Computational Software; Molecular Modeling; Visualization. [www.q-chem.com](http://www.q-chem.com). Retrieved 2020-01-22.
33. Zhao, Y. & Truhlar, D. G. How Well Can New-Generation Density Functionals Describe the Energetics of Bond-Dissociation Reactions Producing Radicals? *J. Phys. Chem. A* **112**, 1095–1099 (2008).
34. Jacquemin, D. *et al.* On the Performances of the M06 Family of Density Functionals for Electronic Excitation Energies. *J. Chem. Theory Comput.* **6**, 2071–2085 (2010).
35. Zhao, Y. & Truhlar, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **4**, 1849–1868 (2008).

36. Zhao, Y. & Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **41**, 157–167 (2008).
37. Tomasi, J., Mennucci, B., Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **105**, 2999 – 3093 (2005).
38. Uhe, A., Kozuch, S. & Shaik, S. Automatic analysis of computed catalytic cycles. *J. Comput. Chem.* **32**, 978–985 (2011).
39. Kozuch, S. & Shaik, S. How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Acc. Chem. Res.* **44**, 101–110 (2011).
40. Kozuch, S. & Shaik, S. A Combined Kinetic–Quantum Mechanical Model for Assessment of Catalytic Cycles: Application to Cross-Coupling and Heck Reactions. *J. Am. Chem. Soc.* **128**, 3355–3365 (2006).
41. Berkessel, A., Elfert, S., Etzenbach-Effers, K. & Teles, J. H. Aldehyde Umpolung by N-Heterocyclic Carbenes: NMR Characterization of the Breslow Intermediate in its Keto Form, and a Spiro-Dioxolane as the Resting State of the Catalytic System. *Angew. Chem. Int. Ed.* **49**, 7120–7124 (2010).
42. Paul, M., Neudörfl, J.-M. & Berkessel, A. Breslow Intermediates from a Thiazolin-2-ylidene and Fluorinated Aldehydes: XRD and Solution-Phase NMR Spectroscopic Characterization. *Angew. Chem. Int. Ed.* **58**, 10596–10600 (2019).
43. Mai, B. K., Park, K., Duong, M. P. T. & Kim, Y. Proton Transfer Dependence on Hydrogen-Bonding of Solvent to the Water Wire: A Theoretical Study. *J. Phys. Chem. B* **117**, 307–315 (2013).
44. Pérez-Soto, R., Besora, M. & Maseras, F. The Challenge of Reproducing with Calculations Raw Experimental Kinetic Data for an Organic Reaction. *Org. Lett.* **22**, 2873–2877 (2020).
45. Essafri, I. & Ghoufi, A. Microstructure of nonideal methanol binary liquid mixtures. *Phys. Rev. E* **99**, 062607 (2019).

46. Mhanna, R., Lefort, R., Noirez, L. & Morineau, D. Microstructure and concentration fluctuations in alcohol–Toluene and alcohol–Cyclohexane binary liquids: A small angle neutron scattering study. *J. Mol. Liq.* **218**, 198–207 (2016).